

---

# NEAR-OPTIMAL ANALYSIS OF UNIVARIATE MOMENT BOUNDS FOR POLYNOMIAL OPTIMIZATION

---

Monique Laurent <sup>\*</sup>

Lucas Slot <sup>†</sup>

January 31, 2020

## ABSTRACT

We consider a recent hierarchy of upper approximations proposed by Lasserre (arXiv:1907.097784, 2019) for the minimization of a polynomial  $f$  over a compact set  $K \subseteq \mathbb{R}^n$ . This hierarchy relies on using the push-forward measure of the Lebesgue measure on  $K$  by the polynomial  $f$  and involves univariate sums of squares of polynomials with growing degrees  $2r$ . Hence it is weaker, but cheaper to compute, than an earlier hierarchy by Lasserre (SIAM Journal on Optimization 21(3), 864–885, 2011), which uses multivariate sums of squares. We show that this new hierarchy converges to the global minimum of  $f$  at a rate in  $O(\log^2 r/r^2)$  whenever  $K$  satisfies a mild geometric condition, which holds, e.g., for convex bodies. As an application this rate of convergence also applies to the stronger hierarchy based on multivariate sums of squares, which extends earlier convergence results to a wider class of compact sets. Furthermore, we show that our analysis is near-optimal by proving a lower bound on the convergence rate in  $\Omega(1/r^2)$  for a class of polynomials on  $K = [-1, 1]$ , obtained by exploiting a connection to orthogonal polynomials.

**Keywords** polynomial optimization · sum-of-squares polynomial · Lasserre hierarchy · push-forward measure · semidefinite programming · needle polynomial

**AMS subject classification** 90C22; 90C26; 90C30

## 1 Introduction

Consider the problem of finding the minimum value taken by an  $n$ -variate polynomial  $f \in \mathbb{R}[x]$  over a compact set  $K \subseteq \mathbb{R}^n$ , i.e., computing the parameter:

$$f_{\min} = \min_{x \in K} f(x). \quad (1)$$

Throughout we also set  $f_{\max} = \max_{x \in K} f(x)$ . Computing the parameter  $f_{\min}$  (or  $f_{\max}$ ) is a hard problem in general, including for instance the maximum stable set problem as a special case. For a general reference on polynomial optimization and its applications, we refer, e.g., to [9, 12].

If we fix a Borel measure  $\lambda$  with support  $K$ , problem (1) may be reformulated as minimizing the integral  $\int_K f(x)\sigma(x)d\lambda(x)$  over all sum-of-squares polynomials  $\sigma \in \Sigma[x]$  that provide a probability density on  $K$  with respect to the measure  $\lambda$ . By bounding the degree of  $\sigma$ , we obtain the following hierarchy of upper bounds on  $f_{\min}$  proposed by Lasserre [11]:

$$f_{\min} \leq f^{(r)} := \min \left\{ \int_K f(x)\sigma(x)d\lambda(x) : \sigma \in \Sigma[x]_r, \int_K \sigma(x)d\lambda(x) = 1 \right\}. \quad (2)$$

Here  $\Sigma[x]$  denotes the set of polynomials that can be written as a sum of squares of polynomials and we set  $\Sigma[x]_r = \Sigma[x] \cap \mathbb{R}[x]_{2r}$ . Since sums of squares of polynomials can be expressed using semidefinite programming, for any fixed  $r \in \mathbb{N}$  the parameter  $f^{(r)}$  can be computed efficiently by semidefinite programming or, even simpler, as the smallest eigenvalue of an appropriate matrix of size  $\binom{n+r}{r}$  ([11], see also [6]).

<sup>\*</sup>Centrum Wiskunde & Informatica (CWI), Amsterdam and Tilburg University, monique.laurent@cwi.nl

<sup>†</sup>Centrum Wiskunde & Informatica (CWI), Amsterdam, lucas.slot@cwi.nl

Recently, Lasserre [13] introduced new, weaker but more economical, upper bounds on  $f_{\min}$  that are based on a *univariate* approach to the problem. For this purpose, he considers the *push-forward measure*  $\lambda_f$  of  $\lambda$  by  $f$ , which is defined by

$$\lambda_f(B) = \lambda(f^{-1}(B)) \quad \text{for any Borel set } B \subseteq \mathbb{R}. \quad (3)$$

Note that for any measurable function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we thus have

$$\int_{f(K)} g(t) d\lambda_f(t) = \int_K g(f(x)) d\lambda(x). \quad (4)$$

We then can define the following hierarchy of upper bounds on  $f_{\min}$ :

$$\begin{aligned} f_{\min} &\leq f_{\text{pfm}}^{(r)} := \min \left\{ \int_{f(K)} ts(t) d\lambda_f(t) : s \in \Sigma[t]_r, \int_{f(K)} s(t) d\lambda_f(t) = 1 \right\} \\ &= \min \left\{ \int_K f(x)s(f(x)) d\lambda(x) : s \in \Sigma[t]_r, \int_K s(f(x)) d\lambda(x) = 1 \right\}. \end{aligned} \quad (5)$$

The difference with the parameter  $f^{(r)}$  is that we now restrict the search to *univariate* sums of squares  $s \in \Sigma[t]_r$ , which we then evaluate at the polynomial  $f$ , leading to the multivariate sum of squares  $\sigma_{\text{pfm}} := s \circ f \in \Sigma[x]_{rd}$  if  $f$  has degree  $d$ . Therefore we have the inequality

$$f_{\min} \leq f^{(rd)} \leq f_{\text{pfm}}^{(r)}. \quad (6)$$

Again, the parameter  $f_{\text{pfm}}^{(r)}$  can be computed efficiently for any fixed  $r$ . But now it can be computed as the smallest eigenvalue of an appropriate matrix of much smaller size  $r + 1$  (see (8) below). Asymptotic convergence of the parameters  $f_{\text{pfm}}^{(r)}$  to  $f_{\min}$  is shown in [13], but no quantitative results are given there. In this paper, we are interested in analyzing the convergence rate of the parameters  $f_{\text{pfm}}^{(r)}$  to the global minimum  $f_{\min}$  in terms of the degree  $r$ .

## 1.1 Previous work

In what follows we always consider for  $\lambda$  the Lebesgue measure on  $K$  (unless specified otherwise). Several results exist on the convergence rate of the parameters  $f^{(r)}$  to the global minimum  $f_{\min}$ , depending on the set  $K$ . The best rates in  $O(1/r^2)$  were shown in [6, 7, 14] when  $K$  belongs to special classes of convex bodies, including the hypercube  $[-1, 1]^n$ , the ball  $B^n$ , the sphere  $S^{n-1}$ , the standard simplex  $\Delta^n$  and compact sets that are locally ‘ball-like’. Furthermore, it was shown in [6] that this analysis is best possible in general (already for  $K = [-1, 1]$  and  $f(x) = x$ ). The starting point for each of these results is a connection between the parameters  $f^{(r)}$  and the smallest roots of certain orthogonal polynomials (see [6, Section 2] and the short recap below).

In [14, Theorems 10-11], a rate in  $O(\log^2 r/r^2)$  was shown for general convex bodies  $K$ , as well as a rate in  $O(\log r/r)$  for general compact sets  $K$  that satisfy a minor geometric condition (see Assumption 1 below). Here, the analysis relies on constructing explicit sum-of-squares densities that approximate well the Dirac delta function at a global minimizer of  $f$ , making use of the so-called ‘needle’ polynomials from [8]. An improved rate in  $O(\log^k r/r^k)$  was shown in [14, Theorem 14] when the partial derivatives of  $f$  up to degree  $k - 1$  vanish at one of its global minimizers on  $K$ .

When  $K$  is a convex body, a convergence rate in  $O(1/r)$  had been shown earlier in [5], by exploiting a link to simulated annealing. There the authors considered sum-of-squares densities of (roughly) the form  $\sigma = s \circ f$ , where  $s(t) = \sum_{k=0}^{2r} (-t/T)^k/k! \in \Sigma[t]_r$  is the truncated Taylor expansion of the exponential  $e^{-t/T}$ . Hence this specific choice of  $s$  (or  $\sigma$ ) provides an upper bound not only for the parameter  $f^{(rd)}$  (as exploited in [5]) but also for the parameter  $f_{\text{pfm}}^{(r)}$  and thus the result of [5] gives directly  $f_{\text{pfm}}^{(r)} - f_{\min} = O(1/r)$  when  $K$  is a convex body.

The result above gives a first quantitative analysis of the parameters  $f_{\text{pfm}}^{(r)}$  for convex bodies. In this paper we improve this convergence analysis and we extend it to a larger class of compact sets.

## 1.2 New results

The main contribution of this paper is the following bound on the convergence rate of the parameter  $f_{\text{pfm}}^{(r)}$  that holds whenever  $K$  satisfies a minor geometric condition.

**Theorem 1.** *Let  $K \subseteq \mathbb{R}^n$  be a compact connected set satisfying Assumption 1 below. Then we have*

$$f_{\text{pfm}}^{(r)} - f_{\min} = O(\log^2 r/r^2).$$

In view of (6), we immediately get the following corollary, extending the rate in  $O(\log^2 r/r^2)$ , shown in [14] for convex bodies, to all connected compact sets  $K$  satisfying Assumption 1.

**Corollary 1.** *Let  $K \subseteq \mathbb{R}^n$  be a compact connected set satisfying Assumption 1. Then we have*

$$f^{(r)} - f_{\min} = O(\log^2 r/r^2).$$

In light of the following special case of [6, Corollary 3.2] our result on the convergence rate of  $f_{\text{pfm}}^{(r)}$  is best possible in general, up to the log-factor.

**Theorem 2** ([6]). *Let  $K = [-1, 1]$  and let  $f(x) = x$ . Then  $f^{(r)} = -1 + \Theta(1/r^2)$ . As a direct consequence, we have  $f_{\text{pfm}}^{(r)} (= f^{(r)}) = -1 + \Omega(1/r^2)$ .*

As an additional result, we extend the lower bound  $\Omega(1/r^2)$  on the error range  $f_{\text{pfm}}^{(r)} - f_{\min}$  to the class of functions  $f(x) = x^{2k}$  with integer  $k \geq 1$ .

**Theorem 3.** *Let  $K = [-1, 1]$  and let  $f(x) = x^{2k}$  for  $k \geq 1$  integer. Then we have  $f_{\text{pfm}}^{(r)} = \Omega(1/r^2)$ .*

Combining Theorem 3 with the fact that  $f^{(r)} = O(\log^{2k} r/r^{2k})$  when  $f(x) = x^{2k}$  (using [14, Theorem 14]), we thus show a large separation between the asymptotic quality of the bounds  $f^{(r)}$  and  $f_{\text{pfm}}^{(r)}$  for this class of functions.

### 1.3 Approach and discussion

As already mentioned above, a crucial ingredient in the analysis of the parameters  $f^{(r)}$  for special compact sets like the hypercube  $[-1, 1]^n$ , the ball, the sphere, or the simplex, is the analysis in the univariate case when  $K = [-1, 1]$  (equipped with the Lebesgue measure or more generally allowing a weight of Jacobi type) and the special polynomial  $f(t) = t$ . Let  $\{p_i \in \mathbb{R}[t] : i \in \mathbb{N}\}$  be the (unique) orthonormal basis of  $\mathbb{R}[t]$  with respect to the inner product  $\langle \cdot, \cdot \rangle_\lambda$  given by

$$\langle p, q \rangle_\lambda = \int_K pq d\lambda \quad \text{for } p, q \in \mathbb{R}[t]. \quad (7)$$

Then, as is shown in [6], the parameter  $f^{(r)}$  coincides with the smallest eigenvalue of the (truncated) *moment matrix*  $M_{\lambda, r}$  of  $\lambda$ , which is defined as

$$M_{\lambda, r} := \left( \int_K t p_i p_j d\lambda \right)_{i, j=0}^r. \quad (8)$$

A classical result on orthogonal polynomials (cf., e.g., [15]) shows that the eigenvalues of  $M_{\lambda, r}$  are given by the roots of  $p_{r+1}$ . Hence, the parameter  $f^{(r)}$  is equal to the smallest root of  $p_{r+1}$ , the asymptotic behaviour of which is well understood and known to be in  $-1 + \Theta(1/r^2)$  when  $\lambda$  is a measure of Jacobi type ([6], see also Lemma 2 below).

Recall that  $\lambda_f$  is the push-forward measure of  $\lambda$  by  $f$ , as defined in (3), and  $f(K) = [f_{\min}, f_{\max}]$  since we assume  $K$  is compact and connected. Let  $\{p_{f, i} : i \in \mathbb{N}\}$  denote the orthonormal basis of  $\mathbb{R}[t]$  with respect to the inner product  $\langle \cdot, \cdot \rangle_{\lambda_f}$  on the interval  $[f_{\min}, f_{\max}]$ . In view of the above discussion, if we use the first (univariate) formulation of  $f_{\text{pfm}}^{(r)}$  in (5), we can immediately conclude that  $f_{\text{pfm}}^{(r)}$  is equal to the smallest eigenvalue of the matrix

$$M_{\lambda_f, r} := \left( \int_{f_{\min}}^{f_{\max}} t p_{f, i} p_{f, j} d\lambda_f \right)_{i, j=0}^r,$$

and also to the smallest root of the orthogonal polynomial  $p_{f, r+1}$ . However it is not clear how to exploit this connection in order to gain information about the convergence rate of the parameters  $f_{\text{pfm}}^{(r)}$  since the orthogonal polynomials  $p_{f, i}$  are not known explicitly in general.

In this paper, we will go back to the idea of trying to find a good sum-of-squares polynomial approximation of the Dirac delta function. As in [14], we make use of the needle polynomials from [8] for this purpose. The difference with the approach in [14] is that we now work on the interval  $[f_{\min}, f_{\max}]$ , and need an approximation of the Dirac delta function centered at  $f_{\min}$ , which is on the boundary of this interval. As is already noted in [8], this special setting allows for better approximations than would be available in general.

### 1.4 Outline

The rest of the paper is organized as follows. In Section 2 we give a proof of Theorem 1. Then, in Section 3, we prove Theorem 3. Finally, we provide some numerical examples that illustrate the practical behaviour of the bounds  $f^{(r)}$  and  $f_{\text{pfm}}^{(r)}$  in Section 4.

## 2 Convergence analysis for the new hierarchy

We first state the precise geometric condition alluded to in Theorem 1.

**Assumption 1.** *There exist positive constants  $\epsilon_K, \eta_K > 0$  such that, for all  $x \in K$  and  $0 < \delta \leq \epsilon_K$ , we have*

$$\text{vol}(K \cap B_\delta^n(x)) \geq \eta_K \delta^n \text{vol}(B^n). \quad (9)$$

Here, for any  $\rho > 0$  and  $x \in \mathbb{R}^n$ ,  $B_\rho^n(x)$  is the Euclidean ball centered at  $x$  with radius  $\rho$  and  $B^n = B_1^n(0)$ .

Assumption 1 was introduced in [4], where it was used to give the first error analysis in  $O(1/\sqrt{r})$  for the bounds  $f^{(r)}$ . This condition on the set  $K$  is rather mild and it is satisfied, e.g., when  $K$  is a convex body, or more generally when  $K$  satisfies an interior cone condition, or when  $K$  is star-shaped with respect to a ball (see [4] for a more complete discussion).

We show the following restatement of Theorem 1.

**Theorem 4.** *Assume  $K$  is connected compact and satisfies the above geometric condition (9). Then there exists a constant  $C$  (depending only on  $n$ , the Lipschitz constant of  $f$  and  $K$ ) such that*

$$f_{\text{pfm}}^{(r)} - f_{\min} \leq C \frac{\log^2 r}{r^2} (f_{\max} - f_{\min}) \quad \text{for all large } r.$$

The rest of this section is devoted to the proof of Theorem 4. We will make the following assumptions in order to simplify notation in our arguments. Let  $a$  be a global minimizer of  $f$  in  $K$ . After applying a suitable translation (replacing  $K$  by  $K - a$  and the polynomial  $f$  by the polynomial  $x \mapsto f(x - a)$ ), we may assume that  $a = 0$ , that is, we may assume that the global minimum of  $f$  over  $K$  is attained at the origin. Furthermore, it suffices to work with the rescaled polynomial

$$F(x) := \frac{f(x) - f_{\min}}{f_{\max} - f_{\min}},$$

which satisfies  $F(K) = [0, 1]$ , with  $F_{\min} = 0$  and  $F_{\max} = 1$ . Indeed, one can easily check that

$$f_{\text{pfm}}^{(r)} - f_{\min} \leq (f_{\max} - f_{\min}) F_{\text{pfm}}^{(r)}.$$

Then, for this polynomial  $F$ , we know that the support of the push-forward measure  $\lambda_F$  is equal to  $[0, 1]$ , and (5) gives

$$\begin{aligned} F_{\text{pfm}}^{(r)} &= \min \left\{ \int_0^1 t s(t) d\lambda_F(t) : s \in \Sigma[t]_r, \int_0^1 s(t) d\lambda_F(t) = 1 \right\} \\ &= \min \left\{ \int_K F(x) s(F(x)) d\lambda(x) : s \in \Sigma[t]_r, \int_K s(F(x)) d\lambda(x) = 1 \right\}. \end{aligned} \quad (10)$$

In order to analyze the bound  $F_{\text{pfm}}^{(r)}$ , we follow a similar strategy to the one employed in [14] to analyze the bound  $F^{(r)}$ . Namely, we construct a univariate sum-of-squares polynomial  $s$  which approximates well the Dirac delta centered at the origin on the interval  $[0, 1]$ , making use of the so-called  $\frac{1}{2}$ -needle polynomials from [8].

**Lemma 1** ([8]). *Let  $h \in (0, 1)$  be a scalar and let  $r \in \mathbb{N}$ . Then there exists a univariate polynomial  $\nu_r^h \in \Sigma[t]_{2r}$  satisfying the following properties:*

$$\begin{aligned} \nu_r^h(0) &= 1, \\ 0 \leq \nu_r^h(t) &\leq 1 \quad \text{for all } t \in [0, 1], \\ \nu_r^h(t) &\leq 4e^{-\frac{1}{2}r\sqrt{h}} \quad \text{for all } t \in [h, 1]. \end{aligned} \quad (11)$$

We consider the sum-of-squares polynomial  $s(t) := C\nu_r^h(t)$ , where  $h \in (0, 1)$  will be chosen later, and  $C$  is chosen so that  $s$  is a density on  $[0, 1]$  with respect to the measure  $\lambda_F$ . That is,

$$C = \left( \int_K \nu_r^h(F(x)) d\lambda(x) \right)^{-1}.$$

As  $s$  is a feasible solution to (10), we obtain

$$F_{\text{pfm}}^{(r)} \leq \int_K F(x) s(F(x)) d\lambda(x) = \frac{\int_K F(x) \nu_r^h(F(x)) d\lambda(x)}{\int_K \nu_r^h(F(x)) d\lambda(x)}.$$

Our goal is thus to show that

$$\text{ratio} := \frac{\int_K F(x) \nu_r^h(F(x)) d\lambda(x)}{\int_K \nu_r^h(F(x)) d\lambda(x)} = O\left(\frac{\log^2 r}{r^2}\right). \quad (12)$$

Define the set

$$K_h = \{x \in K : F(x) \leq h\}.$$

We first work out the numerator of (12), which we split into two terms, depending whether we integrate on  $K_h$  or on its complement:

$$\begin{aligned} \int_K F(x) \nu_r^h(F(x)) d\lambda(x) &= \int_{K_h} F(x) \nu_r^h(F(x)) d\lambda(x) + \int_{K \setminus K_h} F(x) \nu_r^h(F(x)) d\lambda(x) \\ &\leq h \int_{K_h} \nu_r^h(F(x)) d\lambda(x) + \int_{K \setminus K_h} \nu_r^h(F(x)) d\lambda(x). \end{aligned}$$

Here we have upper bounded  $F(x)$  by  $h$  on  $K_h$  and by 1 on  $K \setminus K_h$ . On the other hand, we can lower bound the denominator in (12) as follows:

$$\int_K \nu_r^h(F(x)) d\lambda(x) \geq \int_{K_h} \nu_r^h(F(x)) d\lambda(x).$$

Combining the above two inequalities on numerator and denominator we get

$$\text{ratio} \leq h + \frac{\int_{K \setminus K_h} \nu_r^h(F(x)) d\lambda(x)}{\int_{K_h} \nu_r^h(F(x)) d\lambda(x)}.$$

Thus we only need to upper bound the second term above. We first work on the numerator. For any  $x \in K \setminus K_h$  we have  $F(x) > h$  and thus, using (11), we get  $\nu_r^h(F(x)) \leq 4e^{-\frac{1}{2}r\sqrt{h}}$ . This implies

$$\int_{K \setminus K_h} \nu_r^h(F(x)) d\lambda(x) \leq 4e^{-\frac{1}{2}r\sqrt{h}} \lambda(K).$$

Next, we bound the denominator. In [14, Corollary 4], it is observed that

$$\nu_r^h(t) \geq 1 - 32r^2t \geq \frac{1}{2} \quad \text{for all } t \in [0, \frac{1}{64r^2}].$$

Set  $\rho = \frac{1}{64r^2}$ . We will later choose  $h \geq \rho$ , so that  $K_h \supseteq K_\rho := \{x \in K : F(x) \leq \rho\}$  and  $\nu_r^h(F(x)) \geq \frac{1}{2}$  for all  $x \in K_\rho$ . As  $K$  is compact, there exists a Lipschitz constant  $C_F > 0$  such that

$$F(x) \leq C_F \|x\| \quad \text{for all } x \in K. \quad (13)$$

Note that  $K \cap B_{\rho/C_F}^n \subseteq K_\rho$ . By the geometric assumption (9) we have

$$\lambda(K \cap B_{\rho/C_F}^n) \geq \eta_K \left(\frac{\rho}{C_F}\right)^n \lambda(B^n)$$

for all  $r$  large enough such that  $\rho/C_F \leq \epsilon_K$ . We can then lower bound the denominator as follows:

$$\int_{K_h} \nu_r^h(F(x)) d\lambda(x) \geq \int_{K_\rho} \nu_r^h(F(x)) d\lambda(x) \geq \frac{1}{2} \lambda(K_\rho) \geq \frac{1}{2} \lambda(K \cap B_{\rho/C_F}^n) \geq \frac{1}{2} \eta_K \left(\frac{\rho}{C_F}\right)^n \lambda(B^n).$$

Combining the above inequalities, we obtain

$$\text{ratio} \leq h + \frac{e^{-\frac{1}{2}r\sqrt{h}}}{\rho^n} \cdot \frac{8 \cdot \lambda(K) C_F^n}{\eta_K \lambda(B^n)}.$$

If we now select  $h = \left(4(n+1) \frac{\log r}{r}\right)^2$ , we have  $h \geq \rho$  and a straightforward computation shows that

$$\text{ratio} \leq O\left(\frac{\log^2 r}{r^2}\right).$$

Here, the constant in the big O depends on  $n$ ,  $C_F$  and  $\eta_K$ . This concludes the proof of Theorem 4.

### 3 Separation for a special class of polynomials

In this section we consider in more detail the behaviour of the bounds  $f^{(r)}$  and  $f_{\text{pfm}}^{(r)}$  for the class of polynomials  $f(x) = x^{2k}$  (with  $k \geq 1$  integer) on the interval  $K = [-1, 1]$ . Then  $f([-1, 1]) = [0, 1]$  and, by applying (6) to the polynomial  $f(x) = x^{2k}$ , we have the following inequality:

$$0 \leq f^{(2rk)} \leq f_{\text{pfm}}^{(r)} \quad \text{for any } r \geq 1.$$

Note that for any  $i \leq 2k - 1$ , the  $i$ th derivative of  $f$  vanishes at its global minimizer 0 on  $[-1, 1]$ . Using [14, Theorem 14], we therefore have that  $f^{(2rk)} \leq f^{(r)} = O(\log^{2k} r / r^{2k})$ . On the other hand, the convergence rate in  $O(\log^2 r / r^2)$  for  $f_{\text{pfm}}^{(r)}$  shown in Theorem 1 is optimal up to the log-factor. Indeed, we will show here a lower bound for  $f_{\text{pfm}}^{(r)}$  in  $\Omega(1/r^2)$ .

Let  $\lambda_k := \lambda_f$  denote the push-forward measure (3) of the Lebesgue measure on  $[-1, 1]$  by the function  $f(x) = x^{2k}$ , and let  $\{p_{k,i}(t) : i \in \mathbb{N}\} \subseteq \mathbb{R}[t]$  denote the family of orthogonal polynomials that provide an orthonormal basis for  $\mathbb{R}[t]$  w.r.t. the inner product  $\langle \cdot, \cdot \rangle_{\lambda_k}$  (cf. (7)). Then, as shown in [6] and as recalled above, the parameter  $f_{\text{pfm}}^{(r)}$  is equal to the smallest root of the polynomial  $p_{k,r+1}(t)$ . As it turns out, here we can find explicitly the push-forward measure  $\lambda_k$ , which can be shown to be of Jacobi type. Hence, we have information about the corresponding orthogonal polynomials  $p_{k,i}$ , whose extremal roots are well understood. First we introduce the classical Jacobi polynomials (see, e.g., [15] for a general reference).

**Lemma 2.** *Let  $a, b > -1$ . Consider the weight function  $w_{a,b}(x) = (1-x)^a(1+x)^b$  on the interval  $[-1, 1]$  and let  $\{p_i^{a,b}(x) : i \in \mathbb{N}\}$  be the corresponding family of orthogonal polynomials. Then  $p_i^{a,b}$  is known as the degree  $i$  Jacobi polynomial (with parameters  $a, b$ ), and its smallest root  $\xi_i^{a,b}$  satisfies:*

$$\xi_i^{a,b} = -1 + \Theta(1/i^2). \quad (14)$$

*Proof.* A proof of this fact based on results in [2, 3] is given in [6]. □

**Lemma 3.** *For any integrable function  $g$  on  $[-1, 1]$  we have the identity*

$$\int_{-1}^1 g(x^{2k}) dx = \frac{1}{k} \int_0^1 g(t) t^{-1+1/2k} dt.$$

*Hence, the push-forward measure  $\lambda_k$  is given by  $d\lambda_k(t) := \frac{1}{k} t^{-1+1/2k} dt$  for  $t \in [0, 1]$ .*

*Proof.* It suffices to show the first claim, which follows by making a change of variables  $t = x^{2k}$  so that we get

$$\int_{-1}^1 g(x^{2k}) dx = 2 \int_0^1 g(x^{2k}) dx = 2 \int_0^1 g(t) \frac{t^{-1+1/2k}}{2k} dt = \frac{1}{k} \int_0^1 g(t) t^{-1+1/2k} dt.$$

□

*Proof of Theorem 3.* By applying the change of variables  $x = 2t - 1$ , we see that the Jacobi type measure  $(1-x)^a(1+x)^b dx$  on  $[-1, 1]$  corresponds to the measure  $2^{a+b}(1-t)^a t^b dt$  on  $[0, 1]$  and that (up to scaling) the orthogonal polynomials for the latter measure on  $[0, 1]$  are given by  $t \mapsto p_i^{a,b}(2t - 1)$  for  $i \in \mathbb{N}$ .

If we set  $a = 0$  and  $b = -1 + 1/2k$ , then the measure obtained in this way on  $[0, 1]$  is precisely the push-forward measure  $\lambda_k$  (see Lemma 3). Hence, we can conclude that (up to scaling) the orthogonal polynomials  $p_{k,i}$  for  $\lambda_k$  on  $[0, 1]$  are given by  $p_{k,i}(t) = p_i^{a,b}(2t - 1)$  for each  $i \in \mathbb{N}$ . Therefore, the smallest root of  $p_{k,r+1}(t)$  is equal to  $(\xi_{r+1}^{a,b} + 1)/2 = \Theta(1/r^2)$  by (14). In particular, we can conclude that  $f_{\text{pfm}}^{(r)} = \Omega(1/r^2)$  for any  $k \geq 1$ . □

### 4 Numerical examples

In this section, we illustrate the practical behaviour of the bounds  $f_{\text{pfm}}^{(r)}$  and  $f^{(r)}$  using some numerical examples.

**Comparison of  $f_{\text{pfm}}^{(r)}$  and  $f^{(r)}$  for polynomial test functions.** First, we consider the polynomial test functions listed in Table 1. These are all well-known in optimization, and were already used to test the behaviour of the bounds  $f^{(r)}$

Name	Formula	$f_{\min}$
Booth	$f_{\text{bo}}(x) = (10x_1 + 20x_2 - 7)^2 + (20x_1 + 10x_2 - 5)^2$	$f_{\text{bo}}(\frac{1}{10}, \frac{3}{10}) = 0$
Matyas	$f_{\text{ma}}(x) = 26(x_1^2 + x_2^2) - 48x_1x_2$	$f_{\text{ma}}(0, 0) = 0$
Camel	$f_{\text{ca}}(x) = 50x_1^2 - \frac{2625}{4}x_1^4 + \frac{15625}{6}x_1^6 + 25x_1x_2 + 25x_2^2$	$f_{\text{ca}}(0, 0) = 0$
Motzkin	$f_{\text{mo}}(x) = 64x_1^4x_2^2 + 64x_1^2x_2^4 - 48x_1^2x_2^2 + 1$	$f_{\text{mo}}(\pm\frac{1}{2}, \pm\frac{1}{2}) = 0$

Table 1: Polynomial test functions. In each case,  $f_{\min}$  is the global minimum of  $f$  on  $[-1, 1]^2$ .

in [4, 14]. We compare the bounds  $f^{(r)}$  and  $f_{\text{pfm}}^{(r)}$  directly for  $f \in \{f_{\text{bo}}, f_{\text{ma}}, f_{\text{ca}}, f_{\text{mo}}\}$ , computed for the unit box  $[-1, 1]^2$  and the unit ball  $B^2$ . For  $1 \leq r \leq 20$ , we compute the values of the fraction:

$$\rho_r(f) := \frac{f_{\text{pfm}}^{(r)} - f_{\min}}{f^{(r)} - f_{\min}}.$$

So, values of  $\rho_r(f)$  smaller than 1 indicate good performance of the bounds  $f_{\text{pfm}}^{(r)}$  in comparison to  $f^{(r)}$ . The results can be found in Figure 3. Remarkably, it appears that the performance of the bound  $f_{\text{pfm}}^{(r)}$  is comparable to (or better than) the performance of  $f^{(r)}$  in each instance, except for the Camel function. Additionally, we note that the performance of  $f_{\text{pfm}}^{(r)}$  for the Motzkin polynomial is comparatively much better on the unit ball than on the unit box. Figure 1 shows a plot of the Camel function, as well as the sum-of-squares densities corresponding to  $f^{(6)}$  and  $f_{\text{pfm}}^{(6)}$  on the unit box. Note that while the density corresponding to  $f^{(6)}$  resembles the Dirac delta function centered at the global minimizer  $(0, 0)$  of the Camel function, the density corresponding to  $f_{\text{pfm}}^{(6)}$  instead mirrors the Camel function itself.

**Comparison of  $f_{\text{pfm}}^{(r)}$  and  $f^{(r)}$  for the special class of polynomials  $f(x) = x^{2k}$ .** Next, we consider the polynomials  $f(x) = x^{2k}$  for  $k \geq 1$  on the interval  $[-1, 1]$ , which were treated in Section 3. In Figure 4, the values of  $\rho_r(f)$  are shown for  $1 \leq r \leq 20$  and  $1 \leq k \leq 5$ . It can be seen that the performance of  $f_{\text{pfm}}^{(r)}$  is comparable to the performance of  $f^{(r)}$  for  $k = 1$  (indeed, in this case we have  $f_{\text{pfm}}^{(r)} = f^{(2r)}$ ), but it is much worse for  $k > 1$ , which matches our earlier findings (Theorem 3). In Figure 2, the optimal sum-of-squares densities  $\sigma$  (corresponding to  $f^{(r)}$ ) and  $\sigma_{\text{pfm}}$  (corresponding to  $f_{\text{pfm}}^{(r)}$ ) are depicted for  $k = 1, 3, 5$  and  $r = 6$ . Note that while the density  $\sigma$  changes very little as we increase  $k$ , the density  $\sigma_{\text{pfm}}$  grows increasingly ‘flat’ around the minimizer 0 of  $f$  (mirroring the behavior of  $f$  itself). As such, the density  $\sigma_{\text{pfm}}$  is a comparatively much worse approximation of the Dirac delta function centered at 0 than  $\sigma$ . Note also that in this instance  $f^{(r)} = f^{(r+1)}$  for even  $r$ , explaining the ‘zig-zagging’ behaviour of the ratio  $\rho_r(f)$ .

## 5 Conclusions

We have shown a convergence rate in  $O(\log^2 r/r^2)$  for the approximations  $f_{\text{pfm}}^{(r)}$  of the minimum of a polynomial  $f$  over a compact connected set  $K$  satisfying the minor geometric assumption (9). Furthermore, we have shown that this analysis is near-optimal, in the sense that the asymptotic behaviour of the error range  $f_{\text{pfm}}^{(r)} - f_{\min}$  is in  $O(\log^2 r/r^2)$  in general and in  $\Omega(1/r^2)$  for an infinite class of polynomials.

This latter result shows that although the worst-case guarantees on the convergence of the bounds  $f^{(r)}$  and  $f_{\text{pfm}}^{(r)}$  are very similar, a large separation may exist for certain polynomials (e.g., when  $f(x) = x^{2k}$ ). Of course, it should be noted that the parameter  $f_{\text{pfm}}^{(r)}$  can be obtained via a much smaller eigenvalue computation than the parameter  $f^{(r)}$ , namely by computing the smallest eigenvalue of a matrix of size  $r + 1$  for the latter in comparison to a matrix of size  $\binom{n+r}{r}$  for the former.

Lastly, as a surprising consequence of Theorem 1, we are able to extend the bound in  $O(\log^2 r/r^2)$  on the convergence rate of  $f^{(r)}$  to all compact connected sets  $K$  satisfying the geometric condition (9), whereas it was previously only known for convex bodies [14]. In this sense, the arguments of Section 2 can be seen as a refinement (and simplification) of the ones given in [14].

As said above, the analysis in this paper is near-optimal: we can show an upper bound in  $O(\log^2 r/r^2)$  and a lower bound in  $\Omega(1/r^2)$  for a certain class of polynomials. Deciding what is the right regime and whether the log-factor can be avoided in the convergence analysis is the main research question left open by this work.

The log-factor arises from our analysis technique, based on using polynomial approximation by the needle polynomials. We had to use this analysis technique since the behaviour of the orthogonal polynomials for the push-forward

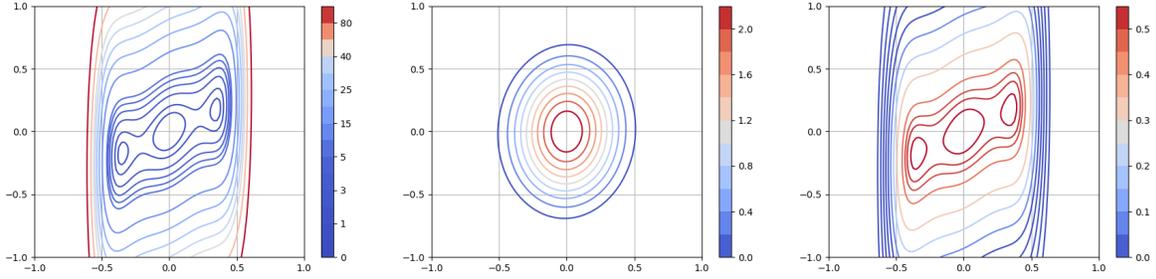


Figure 1: The Camel function (left) and its sum-of-squares densities corresponding to  $f^{(6)}$  (middle) and  $f_{\text{pfm}}^{(6)}$  (right) on the unit box.

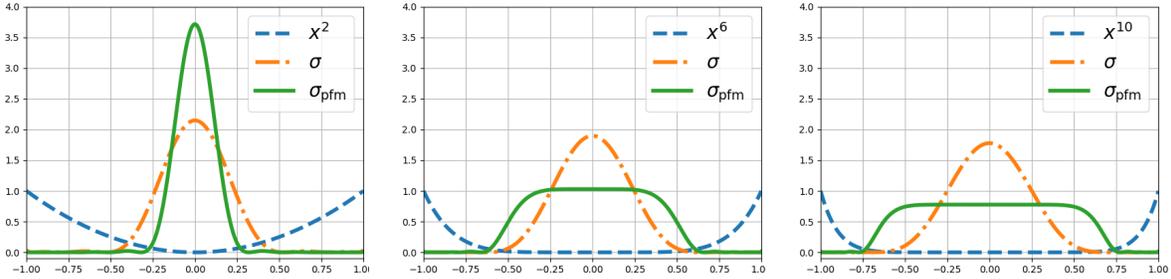


Figure 2: The functions  $f(x) = x^{2k}$  and their sum-of-squares densities corresponding to  $f^{(6)}$  and  $f_{\text{pfm}}^{(6)}$  on the interval  $[-1, 1]$  for  $k = 1$  (left),  $k = 3$  (middle) and  $k = 5$  (right).

measure  $\lambda_f$  is not known for general  $f$ . On the other hand, our results may be interpreted as giving back some information for general push-forward measures  $\lambda_f$  and their corresponding orthogonal polynomials  $p_{f,i}$  on the interval  $[f_{\min}, f_{\max}]$ . Indeed, what our results imply is that for any polynomial  $f$  and any compact connected  $K$  satisfying (9), the asymptotic behaviour of the smallest root of  $p_{f,i}$  is in  $f_{\min} + O(\log^2 r/r^2)$ .

## Acknowledgments

This work is supported by the Europeans Union’s EU Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie Actions Grant Agreement No 764759 (MINOA).

## References

- [1] F. Dai, Y. Xu. Approximation Theory and Harmonic Analysis on Spheres and Balls. Springer, New York (2013)
- [2] D.K. Dimitrov, G.P. Nikolov. Sharp bounds for the extreme zeros of classical orthogonal polynomials. Journal of Approximation Theory **162**, 1793–1804 (2010)
- [3] K. Driver, K. Jordaan. Bounds for extreme zeros of some classical orthogonal polynomials. Journal of Approximation Theory **164**, 1200–1204 (2012)
- [4] E. de Klerk, M. Laurent, Z. Sun. Convergence analysis for Lasserre’s measure-based hierarchy of upper bounds for polynomial optimization, Math. Program. Ser. A **162**(1), 363–392 (2017)
- [5] E. de Klerk, M. Laurent. Comparison of Lasserre’s measure-based bounds for polynomial optimization to bounds obtained by simulated annealing. Mathematics of Operations Research **43**, 1317–1325 (2018)
- [6] E. de Klerk, M. Laurent. Worst-case examples for Lasserre’s measure-based hierarchy for polynomial optimization on the hypercube. arXiv:1804.05524, to appear in Mathematics of Operations Research (2019)

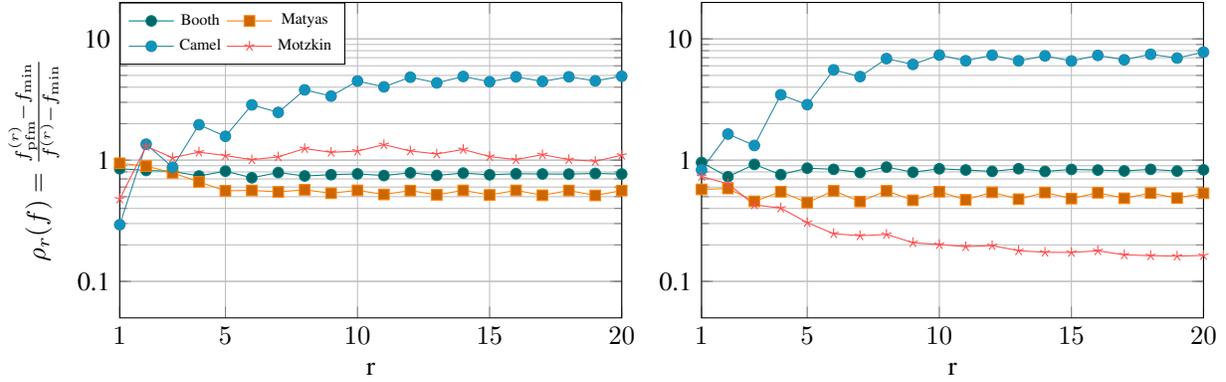


Figure 3: Comparison of the bounds  $f^{(r)}$  and  $f_{\text{pfm}}^{(r)}$  for the first four functions in Table 1, computed on the unit box (left) and unit ball (right).

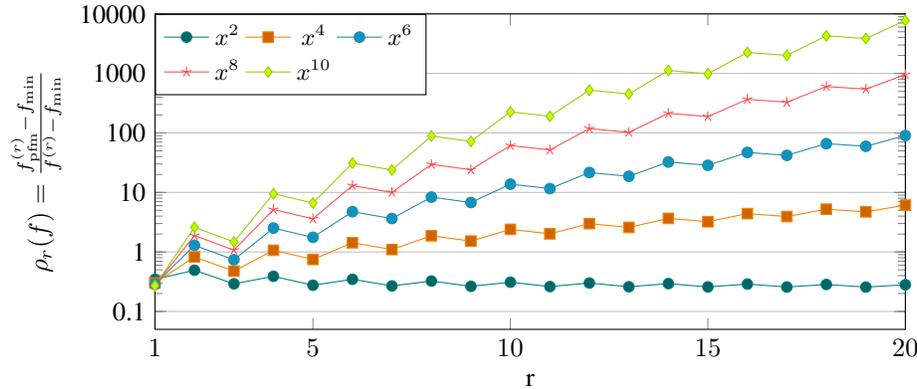


Figure 4: Comparison of the bounds  $f^{(r)}$  and  $f_{\text{pfm}}^{(r)}$  for functions of the form  $f(x) = x^{2k}$  on the interval  $[-1, 1]$ .

- [7] E. de Klerk, M. Laurent. Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. *Mathematical Programming* (2020). <https://doi.org/10.1007/s10107-019-01465-1>
- [8] A. Kroó. Multivariate “needle” polynomials with application to norming sets and cubature formulas. *Acta Mathematica Hungarica* **147**(1), 46–72 (2015)
- [9] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*. Imperial College Press, London (2009)
- [10] J.-B. Lasserre. Bounding the support of a measure from its marginal moments. *Proceedings of the AMS* **139**, 3375–3382 (2011)
- [11] J.-B. Lasserre. A new look at nonnegativity on closed sets and polynomial optimization. *SIAM Journal on Optimization* **21**(3), 864–885 (2011).
- [12] J.-B. Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization* (Cambridge Texts in Applied Mathematics). Cambridge University Press, Cambridge (2015)
- [13] J.-B. Lasserre. Connecting optimization with spectral analysis of tri-diagonal Hankel matrices. arXiv:1907.097784 (2019)
- [14] L. Slot, M. Laurent. Improved convergence analysis of Lasserre’s measure-based upper bounds for polynomial minimization on compact sets. *Mathematical Programming* (2020). <https://doi.org/10.1007/s10107-020-01468-3>
- [15] G. Szegő. *Orthogonal Polynomials*. American Mathematical Society Colloquium Publications, Providence, (1975)